Combining deep learning and shape priors for bi-ventricular segmentation of volumetric cardiac magnetic resonance images

Duan, Jinming; Schlemper, Jo; Bai, Wenjia; Dawes, Timothy J.W.; Bello, Ghalib; Biffi, Carlo; Doumou, Georgia; De Marvao, Antonio; O'Regan, Declan P.; Rueckert, Daniel

DOI:
10.1007/978-3-030-04747-4_24

Combining deep learning and shape priors for bi-ventricular segmentation of volumetric cardiac MR images

Jinming Duan[1,2(✉)], Jo Schlemper[1], Wenjia Bai[1], Timothy J W Dawes[2], Ghalib Bello[2], Carlo Biffi[1,2], Georgia Doumoud[2], Antonio De Marvao[2], Declan O'Regan[2], Daniel Rueckert[1]

[1]Biomedical Image Analysis Group, Imperial College London, London, UK
[2]MRC London Institute of Medical Sciences, Imperial College London, London, UK
j.duan@imperial.ac.uk

Abstract. Deep learning approaches such as convolutional neural networks (CNN) have achieved state-of-the-art performance in cardiac MR (CMR) image segmentation. However, it is non-trivial to introduce shape prior knowledge to CNN-based approaches. In this paper, we combine a CNN-based method with image registration to develop a shape-based bi-ventricular segmentation tool for short-axis CMR volumetric images. The method first employs a fully convolutional network (FCN) to learn the segmentation task from manually labelled ground truth CMR volumes. However, due to existing image artefacts in the training dataset, the resulting FCN segmentation results are often imperfect. As such, we propose a second step to refine the FCN segmentation. This step involves performing a non-rigid registration with multiple high-resolution bi-ventricular atlases, allowing the explicit shape priors to be inferred. We validate the proposed approach on 1831 healthy subjects and 200 subjects with pulmonary hypertension. Numerical experiments on the two datasets demonstrate that our approach is capable of producing accurate, high-resolution and anatomically smooth bi-ventricular models, despite the artefacts in the input CMR volumes.

1 Introduction

Cardiac MR (CMR) imaging is a non-invasive and non-ionising imaging technique that produces high image quality and excellent soft tissue contrast. Among existing imaging techniques, it has established itself as the gold standard for assessing cardiac chamber volume and mass for a wide range of cardiovascular diseases [1]. CMR imaging techniques, together with semi-automated or automated CMR segmentation algorithms [2,3,4,5,6,7,8,9,10,11], have shown a great impact on studying, understanding and diagnosing cardiovascular diseases. However, there are drawbacks in current CMR segmentation methods.

Anatomically, a human heart is composed of the left ventricle (LV) and the right ventricle (RV). Each ventricle can be subdivided into the cavity region (left ventricular cavity [LVC] and right ventricular cavity [RVC]) and the wall region

(left ventricular wall [LVW] and right ventricular wall [RVW]). Most of the segmentation techniques have only focused on the LVC and LVW [3,4] (or at most the LVC, LVW and RVC [2,5,6,7,8,9,10]). Few studies have attempted a full bi-ventricular segmentation (i.e. LVC+LVW+RVC+RVW) due to the narrow structure of RVW (sometimes less than one millimetre in thickness). This prohibits accurate cardiac assessments involving coupled bi-ventricular cardiac motion. In addition, due to the limitations of standard clinical acquisition protocols, the raw volumetric CMR images acquired often contain several artefacts [12], including intensity inhomogeneity, inter-slice shift (i.e. respiratory motion), large slice thickness, lack of slice coverage, etc. Most existing segmentation methods [3,4,5,6,7,8,9,10,11] deal with CMR volumes directly without taking the artefacts into account. As such, the resulting segmentation inevitably inherit these artefacts. Building an accurate, motion-free, automatically meaningful bi-ventricular segmentation model therefore remains an open problem.

To overcome the aforementioned limitations of current approaches, in this paper we propose a novel approach that addresses the problem of bi-ventricular segmentation of short-axis CMR volumetric images. We make the following three distinct contributions. First, the proposed approach segments an input cardiac volume into LVC, LVW, RVC and RVW. The technique introduced herein is the first one capable of producing a full high-resolution bi-ventricular segmentation in 3D. Second, we introduce anatomical shape prior knowledge (via image registration techniques) to a deep learning approach by using a cohort of high-resolution atlas shapes. As such, the proposed approach is capable to produce an accurate, motion-free and clinically meaningful bi-ventricular segmentation model, despite the existing artefacts in the input volume. Third, we thoroughly assess the effectiveness and robustness of proposed method using two datasets, including high- and low-resolution cardiac volumes from 1831 healthy subjects and 200 pathological subjects, respectively. To our knowledge, this is one of the first CMR segmentation studies utilising datasets of this scale. To quantitatively evaluate our proposed segmentation algorithm, we also develop a method that is able to simulate the artefacts in CMR volumes.

2 Methodology

Fully convolutional network: We treat the problem of predicting segmentation maps as the multi-class classification problem. First, let us formulate the learning problem as follows: we denote the input training dataset by $S = \{(U_i, R_i), i = 1, ..., N_t\}$, where N_t is the number of training data, $U_i = \{u_j^i, j = 1, ..., |U_i|\}$ is the raw input CMR volume, $R_i = \{r_j^i, j = 1, ..., |R_i|\}$, $r_j^i \in \{1, ..., Nr\}$ is the ground truth region labels for volume U_i ($N_r = 5$ representing the LVC, LVW, RVC, RVW and background regions). Note that $|U_i| = |R_i|$ stands for the total number of voxels in a CMR volume. We then define all network layer parameters as $\mathbf{W}$. In a supervised setting, we propose to minimise the following objective function via standard (back-propagation) stochastic gradient descent

(SGD)

$$\mathbf{W}^* = \text{argmin}(L_S(\mathbf{W}) + \alpha L_D(\mathbf{W}) + \beta \|\mathbf{W}\|_F^2), \tag{1}$$

where α and β are weight coefficients balancing the three terms. $L_S(\mathbf{W})$ and $L_D(\mathbf{W})$ are the region associated losses that enable the network to predict segmentation maps. $\|\mathbf{W}\|_F^2$, known as the weight decay term, represents the Frobenius norm on the weights $\mathbf{W}$. This term is used to prevent over-fitting of the network. The training problem is to estimate the parameters $\mathbf{W}$ associated with all the convolutional layers and by minimising (1) the network is able to predict segmentation maps. The definitions of $L_S(\mathbf{W})$ and $L_D(\mathbf{W})$ are given separately as follows:

$$L_S(\mathbf{W}) = -\sum_i \sum_k \sum_{j \in X_k^i} \log P(r_j^i = k | U_i, \mathbf{W}), \tag{2}$$

where i, k and j respectively denote the training sample index, the region label index and the voxel index. X_k^i represents the voxels in training sample i that fall in the region for which the label value is k. $P(r_j^i = k | U_i, \mathbf{W})$ corresponds to the softmax probability estimated by the network for a specific voxel j (subject to the restriction $r_j^i = k$), given the training volume U_i and network weights $\mathbf{W}$. Note that (2) is known as the categorical cross-entropy loss or multi-class logistic loss, in which the summations are carried out over all voxels, labels and training samples.

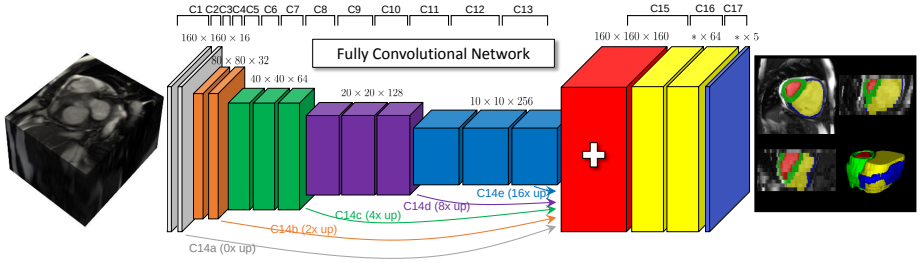

Fig. 1: The architecture of a fully convolutional network with 17 convolutional layers. The network takes the CMR volume as input, applies a branch of convolutions, learns image features from fine to coarse levels, concatenates multi-scale features and finally segments the image into 5 disjoint regions.

Along with (2) for predicting segmentation maps, we use the Dice loss that evaluates spatial overlap with ground truth region labels. More specifically, we use a differentiable approximation of Dice loss, defined as follows:

$$L_D(\mathbf{W}) = -\sum_i \frac{2\sum_k \sum_j 1_{\{r_j^i = k\}} \cdot P(r_j^i = k | U_i, \mathbf{W})}{\sum_k \sum_j \left(1^2_{\{r_j^i = k\}} + P^2(r_j^i = k | U_i, \mathbf{W})\right)}, \tag{3}$$

where $1_{\{\cdot\}}$ is the indicator function and other notations in (3) have the same meanings as those in (2).

In Fig 1, we show the proposed network architecture for automatic CMR segmentation, which is a fully convolutional network (FCN). It is adapted from [10] and similar to the U-net architecture [13]. Batch-normalisation (BN) is used after each convolutional layer, and before a rectified linear unit (ReLU) activation. The last layer is followed by the channel-wise softmax function. In the FCN, input images have pixel dimensions of 192 × 192. Every layer whose label is prefixed with 'C' performs the operation: convolution → BN → ReLU, except C17. The (filter size/stride) is (3×3/1) for layers from C1 to C16, excluding layers C3, C5, C8 and C11 which are (3×3/2). The arrows represent (3×3/1) convolutional layers (C14a−e) followed by a transpose convolutional (up) layer with a factor necessary to achieve feature map volumes with size 160 × 160 × 32, all of which are concatenated into the red feature map volume. Finally, C17 applies a (1×1/1) convolution with a softmax activation, producing the blue feature map volume with a depth 5, corresponding to 5 segmented regions of an input volume.

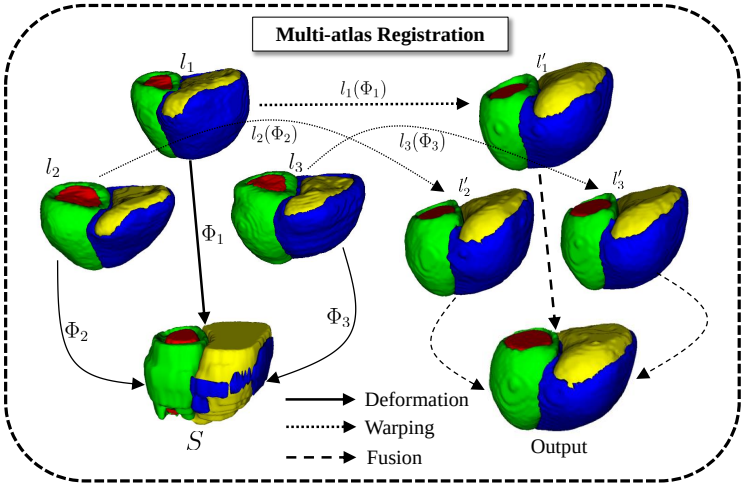

Fig. 2: Incorporation of shape constraints by multi-atlas registration. l_1, l_2 and l_3 are high-resolution bi-ventricular atlas shape models. They are warped to l'_1, l'_2 and l'_3 using the transform between the FCN result S and each atlas. The warped results are finally fused together to generate a smooth high-resolution output.

Multi-atlas registration: As Fig 1 shows, the segmentation produced by FCN is influenced by respiratory motion artefact. Moreover, as the CMR volume are low-resolution in the long-axis, the 3D segmentation model is not smooth. Further, due to the narrow structure of the RVW, the segmentation model is incomplete. By incorporating shape prior knowledge with the following image registration, these drawbacks can be resolved.

Since the correspondences of structures across both target and atlas volumes are explicitly encoded in their segmentations, we only use segmentations for

the following non-rigid registration. Let S and l_n $(n = 1, ..., L)$ be the network segmentation and the nth atlas segmentation (i.e. shape, respectively. Note that here we use a cohort of high-resolution atlases, each of which has an image resolution of $1.25 \times 1.25 \times 2.0$ mm. Let $P_{S,l_n}(i, j)$ be the joint probability of labels i and j in S and l_n, respectively. It is estimated as the number of voxels with label i in S and label j in l_n divided by the total number of voxels in the overlap region of both segmentations. We then maximise the overlap of structures denoted by the same label in both S and l_n by minimising the following objective function

$$\Phi_n^* = \arg\min \mathcal{C}\left(S, l_n(\Phi_n)\right) \tag{4}$$

where Φ_n is the transformation between S and l_n, which is modelled by a free-form deformation (FFD) based on B-splines [14]. $\mathcal{C}(S, l_n) = \sum_{i=1}^{N_r} P_{S,l_n}(i, i)$, representing the label consistency [15]. $\mathcal{C}$ in (4) is a similarity measure of how many labels of all the labels in the atlas segmentation are correctly mapped into the target segmentation. The measure is zero when none of the atlas labels has been correctly mapped into the target segmentation. The measure is one when all reference labels are correctly matched. Gradient descent is then used to minimise the objective function (4). After the optimal Φ_n^* is found, the segmentation in the nth atlas is warped to the target space (i.e. l'_n). The process is repeated several times until after all the pre-selected atlases are warped. Lastly, the resulting label at each voxel in the target volume can be calculated by finding the maximum label of all the warped atlas segmentations at that voxel.

3 Experimental results

For the training and evaluation of the proposed method, we use the UK Digital Heart Project Dataset[1], which is composed of 1831 cine high-resolution CMR volumetric images and the corresponding dense segmentation annotations at the end-diastolic (ED) and end-systolic (ES) frames. These volumes are derived from healthy subjects, scanned at Hammersmith Hospital, Imperial College London using a 3D cine balanced steady-state free precession (b-SSFP) sequence [16] and has a resolution of $1.25 \times 1.25 \times 2$ mm. The high-resolution imaging technique enables us to characterise the cardiac shape in great detail. Moreover, it requires only one single breath-hold of a subject during each scan and thereby does not produce the artefacts (i.e. inter-slice shift, large slice thickness, lack of slice coverage, etc) [16], which are commonly seen when low-resolution imaging techniques [12] are used. Fig 5 *a* shows a long-axis view of a high-resolution volume at the ED frame, and the corresponding ground truth 2D and 3D segmentation labels are given in *c* and *d*, respectively.

To quantitatively study our proposed segmentation algorithm, we develop a method to simulate the artefacts in low-resolution cardiac volumes. Specifically, the original high-resolution volume and its segmentation are first downsampled from $1.25 \times 1.25 \times 2$ mm to $1.25 \times 1.25 \times 10$ mm, as shown in the second

[1] https://digital-heart.org/

row of Fig 5. As evident, after this step the segmentation shape takes on the staircase artefacts as now the downsampled versions have a relatively low long-axis resolution. Moreover, the segmentation around the apical region becomes incomplete due to the lack of slice coverage of the whole heart. We further simulate inter-slice shift artefact by randomly translating each 2D short-axis slice independently. After this step the cardiac volume and its segmentation become misaligned, as shown in the last row of Fig 5. Next, for training the network the low-resolution volume g and its segmentation h are used as inputs. Note that our method is capable of producing a high-resolution smooth segmentation model even through the input volume is like g. Since we have the smooth ground truth c for g, we can quantitatively compare the output of our method with the ground truth c.

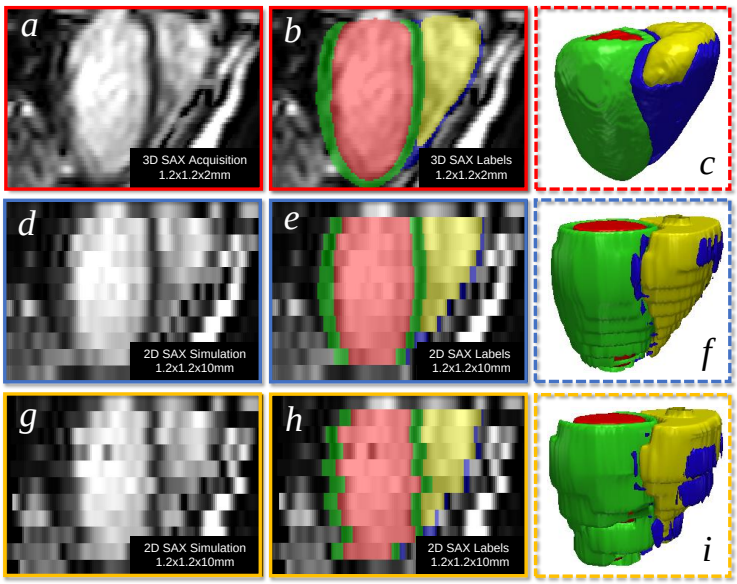

Fig. 3: Simulating cardiac artefacts in real scenarios. 1st row: artefact-free high-resolution cardiac volume and ground truth labels. 2nd row: downsampled versions of volumes in the 1st row. 3rd row: inter-slice shift is added to the downsampled volumes in the 2nd row.

We then randomly split the 1831 dataset into three sets of 1000/600/231. The first two sets are then corrupted with the simulated artefacts introduced above, which are respectively used for training the neural network in Fig 1 and evaluating the proposed segmentation algorithm. The last set remains unchanged and is used as a cohort of high-resolution atlas shapes for refining the network segmentation. Note that we intend to segment a cardiac volume into the left ventricular cavity (LVC), right ventricular cavity (RVC), left ventricular wall (LVW) and right ventricular wall (RVW).

Table 1 reports the Dice metric and Hausdorf distance between automated and manual segmentations, evaluated on the test set of 600 subjects at ED

and ES. The mean Dice values of LVC, LVW and RVC demonstrate a good agreement between automated and manual segmentations for these structures. However, for RVW its mean Dice values at ED and ES are only 0.557 and 0.608. This is due to its thin structure (only two or three voxels in thickness) and the Dice index is more sensitive to errors in this structure. However, in terms of the Hausdorff distance for RVW, the mean value is relatively small. Hence, our method achieves a better performance for all the four structures.

Table 1: The Dice metric and Hausdorf distance (HD) between automated segmentation and manual segmentation for 600 short-axis volumetric images. The mean ± standard deviation are reported at the ED and ES frames.

Region	Dice (ED)	HD in mm (ED)	Dice (ES)	HD in mm (ES)
LVC	0.940±0.024	2.045 ±0.675	0.910±0.028	2.027±0.632
LVW	0.823±0.049	2.394 ±0.841	0.892±0.033	2.431±0.797
RVC	0.914±0.033	3.039 ±1.218	0.901±0.038	2.933±1.253
RVW	0.557±0.121	4.119 ±1.956	0.608±0.123	4.378±2.717

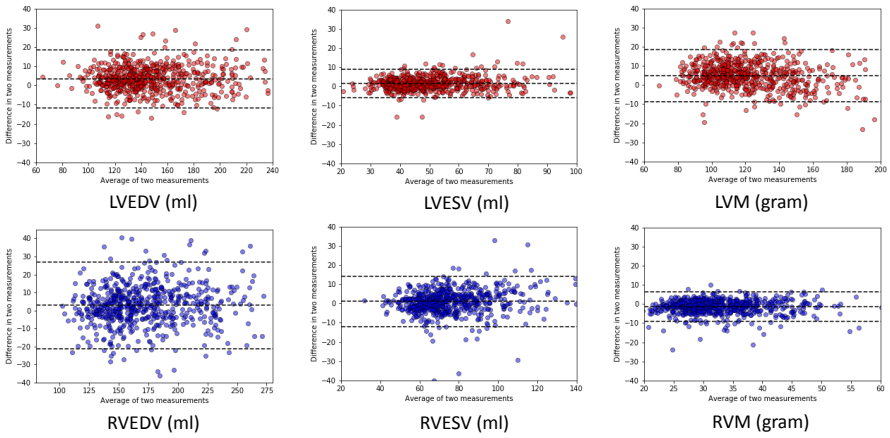

Fig. 4: Bland-Altman plots of clinical measures between automated measurement and manual measurement. The LV end-diastolic volume (LVEDV), end-systolic volume (LVESV), LV myocardial mass (LVM), RV end-diastolic volume (RVEDV), end-systolic volume (RVESV), and RV myocardial mass (RVM) are derived from our segmentation method and the manual segmentation.

To further quantitatively evaluate the proposed method, Fig 4 shows the Bland-Altman plots of the clinical measures. The Bland-Altman plot is commonly used for analysing agreement and bias between two measurements. This figure compares automated measurements to manual measurements on the evaluation set, which shows that the mean difference is centred close to zero, indicating that the automated measurement is almost unbiased relative to the observer. Also, there is no evidence of bias over hearts of difference sizes or volumes. In

particular, RVM shows a very good consistency between the two measurements, validating an accurate segmentation of RVW despite relatively lower Dice values.

Finally, the proposed method was evaluated on a dataset of 200 patients with pulmonary hypertension. Greyscale volumetric images were acquired at low resolution ($1.38 \times 1.38 \times 10$ mm) and segmented into high-resolution smooth 3D models. Results were visually assessed by one clinician with over five years' experience of CMR imaging and judged satisfactory in all cases. In Fig 5, we present an exemplary segmentation of a cardiac volume in a pulmonary hypertension patient. We visually compare the proposed method with the vanilla deep learning method without shape prior knowledge [10]. As the figure shows, the proposed method gives a better 3D phenotype result which is smooth, accurate and artefact-free. This is due to the application of shape prior information. Our method thus outperforms the vanilla FCN in this regard.

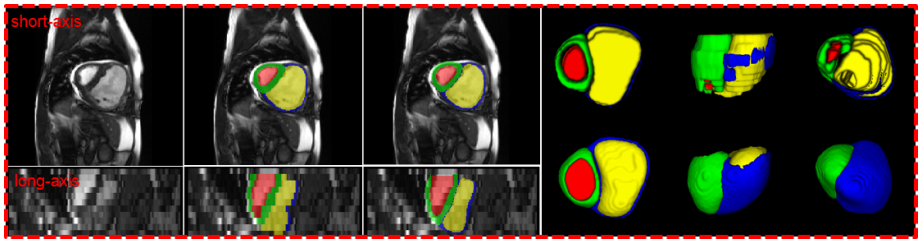

Fig. 5: Visual comparison of segmentation results from the vanilla FCN and the proposed method on a pathological case. 1st column: original short- and long-axis CMR slices. 2nd column: vanilla FCN results. 3rd column: results by the proposed method. Last column: FCN results (top) and our results (bottom). The 5 segmented regions are respectively RVC (yellow), LVC (red), RVW (blue), LVW (green) and background.

4 Conclusion

In this paper, we developed a shape-based CNN-based method for bi-ventricular segmentation of cardiac MR volumetric images. The method first employs a fully convolutional network (FCN) to segment the volume at a low-resolution level. Based on the FCN results, the method then performs the non-rigid registration by using multiple high-resolution atlas shapes, thereby imposing shape constraints explicitly and effectively. Extensive experiments have showed that the method has capability of producing smooth bi-ventricular segmentation results that follow the global anatomical properties of the underlying anatomy, even through the input volumetric images contain several unpleasant artefacts. In addition, we have also showed that the method has a very good generalisation ability for segmentation of pathological cases. Future work will focus on statistical shape analysis using the smooth results produced by the method.

References

1. Ripley, D., Musa, T., Dobson, L., Plein, S., Greenwood, J.: Cardiovascular magnetic resonance imaging: what the general cardiologist should know. Heart (2016) Heartjnl–2015
2. Bai, W., Shi, W., Ledig, C., Rueckert, D.: Multi-atlas segmentation with augmented features for cardiac mr images. Med. Image Anal. **19**(1) (2015) 98–109
3. Nasr-Esfahani, M., Mohrekesh, M., Akbari, M., Soroushmehr, S., Nasr-Esfahani, E., Karimi, N., Samavi, S., Najarian, K.: Left ventricle segmentation in cardiac mr images using fully convolutional network. Arxiv Pre.:1802.07778 (2018)
4. Ngo, T., Lu, Z., Carneiro, G.: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. Med. Image Anal. **35** (2017) 159–171
5. Patravali, J., Jain, S., Chilamkurthy, S.: 2d-3d fully convolutional neural networks for cardiac mr segmentation. Arxiv Pre.:1707.09813 (2017)
6. Baumgartner, C., Koch, L., Pollefeys, M., Konukoglu, E.: An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation. Arxiv Pre.:1709.04496 (2017)
7. Isensee, F., Jaeger, P., Full, P., Wolf, I., Engelhardt, S., Maier-Hein, K.: Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. Arxiv Pre.:1707.00587 (2017)
8. Zheng, Q., Delingette, H., Duchateau, N., Ayache, N.: 3d consistent & robust segmentation of cardiac images by deep learning with spatial propagation. IEEE Trans. Med. Imaging (2018)
9. Tran, P.: A fully convolutional neural network for cardiac segmentation in short-axis mri. Arxiv Pre.:1604.00494 (2016)
10. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A., Aung, N., Lukaschuk, E., Sanghvi, M., et al.: Human-level cmr image analysis with deep fully convolutional networks. Arxiv Pre.:1710.09289 (2017)
11. Duan, J., Schlemper, J., Bai, W., Dawes, T., Bello, G., Doumoud, G., Marvao, A., O'Regan, D., Matthews, P., Rueckert, D.: Deep nested level sets: Fully automated segmentation of cardiac mr images in patients with pulmonary hypertension. In: MICCAI, Springer (2018)
12. Petersen, S., Matthews, P., Francis, J., Robson, M., Zemrak, F., Boubertakh, R., Young, A., Hudson, S., Weale, P., Garratt, S., et al.: Uk biobanks cardiovascular magnetic resonance protocol. J. Cardiovasc. Magn. Reson **18**(1) (2015) 8
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI, Springer (2015) 234–241
14. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. IEEE Trans. Med. Imaging **18**(8) (1999) 712–721
15. Frangi, A., Rueckert, D., Schnabel, J., Niessen, W.: Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. IEEE Trans. Med. Imaging **21**(9) (2002) 1151–1166
16. De Marvao, A., Dawes, T., Shi, W., Minas, C., Keenan, N., Diamond, T., Durighel, G., Montana, G., Rueckert, D., Cook, S., et al.: Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. J. Cardiovasc. Magn. Reson **16**(1) (2014) 16